Environmental Health Perspectives Supplements
Vol. 102 Suppl 1:127-130 1994

Predicting Carcinogenicity by Using Batteries of Dependent Short-Term Tests

by Byung Soo Kim[1] and Barry H. Margolin[2]

Among the various methods for predicting carcinogenicity from a battery of short-term tests (STTs), the carcinogenicity prediction and battery selection (CPBS) procedure is the most prominent. A major assumption of CPBS is that the STTs used in the prediction are conditionally independent. Results of recent National Toxicology Program studies of four commonly used *in vitro* STTs contradict this assumption, thereby necessitating modification of CPBS to accommodate dependencies. This is accomplished via log-linear modeling, which then also yields an important dividend: standard errors for the predicted probabilities of carcinogenicity.

Introduction

Certain classes of carcinogens are poorly detected by individual short-term tests (STTs). Hence, the use of a battery of STTs for detecting carcinogens has been recommended frequently (*1–6*). Several mathematical methods have been presented for prediction of carcinogens from batteries of STTs (*1–3,7,8*). A major assumption of these approaches to prediction is that STTs are conditionally independent and hence that additional STTs will allow detection of additional carcinogens.

The focus here is on the carcinogenicity prediction and battery selection (CPBS)* method, which was developed by Rosenkranz et al. (*2*) and applied to several known databases (*4–6,9*). CPBS employs the Bayes theorem to calculate the conditional probability of carcinogenicity of a chemical given a battery of STTs. As mentioned above, a major assumption of CPBS is that STTs composing the battery are conditionally independent given the carcinogenicity or noncarcinogenicity of the test chemical.

Specifically, suppose one has a battery of r STTs, whose qualitative (positive/negative) results for a given chemical are denoted by $A_1, A_2, \cdots A_r$. Then Rosenkranz et al. (*2*) employed the Bayes theorem to calculate the probability of carcinogenicity given $A_1, \cdots, A_r$ as follows:

$$p(CA \mid A_1 \cdots A_r) = \frac{p(A_1 \cdots A_r \mid CA)\, p(CA)}{p(A_1 \cdots A_r \mid CA)\, p(CA) + p(A_1 \cdots A_r \mid NC)\, p(NC)} \quad (1)$$

They then invoked conditional independence to obtain

$$\frac{\left[\prod_{i=1}^{r} p(A_i \mid CA)\right] p(CA)}{\left[\prod_{i=1}^{r} p(A_i \mid CA)\right] p(CA) + \left[\prod_{j=1}^{r} p(A_j \mid NC)\right] p(NC)} \quad (2)$$

Here CA and NC stand for carcinogenicity and noncarcinogenicity, respectively. When A_i is positive, $p(A_i \mid CA)$ is the sensitivity of the ith STT. When A_i is negative, $p(A_i \mid NC)$ is the specificity of the ith STT. A similar argument can be made regarding $p(NC \mid A_1 \cdots A_r)$, the probability of noncarcinogenicity given $A_1 \cdots A_r$. Equation (*2*) is valid only under the assumption of conditional independence of the r STTs. The statistical independence of several STTs has been documented mostly by showing pairwise independence among those STTs via the Pearson chi-square test. The analysis of published data indicated that about 90% of the pairs of STTs examined showed no evidence of dependence in terms of the chi-square test (*1,10*). This inference, however, is subject to biases of chemical and test selection plus low power to detect dependencies.

In this paper, four widely used STTs are shown to be statistically dependent. The impact of this statistical dependence of the four STTs on carcinogenicity prediction systems is explored in depth, and CPBS is extended to allow for conditionally dependent STTs. This permits examination of the sensitivity of carcinogenicity predictions to the assumption of statistical independence when various possible dependence structures are considered.

Data

In 1984, the National Toxicology Program (NTP) initiated a project with 73 chemicals to develop a database that would permit evaluation of the ability of four of the most commonly used *in vitro* STTs to predict rodent carcinogenicity: the Ames

[1] Department of Applied Statistics, Yonsei University, Seoul, 120-749, South Korea.

[2] Department of Biostatistics, School of Public Health, Lineberger Comprehensive Cancer Center, University of North Carolina, CB 7400, Chapel Hill, NC 27599-7400.

Address reprint requests to B. S. Kim, Department of Applied Statistics, Yonsei University, Seoul, 120-749, South Korea.

* CPBS is a trademark of Case Western Reserve University.

This paper was presented at the International Biostatistics Conference on the Study of Toxicology that was held May 13-25, 1991, in Tokyo, Japan.

Salmonella/microsome mutagenicity assay (SAL), the assays for chromosome aberration (ABS) and sister chromatid exchange (SCE) induction in Chinese hamster ovary cells, and the mouse lymphoma L5178Y cell mutagenesis assay [MLA (*11*)]. That effort differed from previous investigations in two aspects (*11*). First, standard protocols for the four STTs were shown to yield reproducible results in interlaboratory trials with coded chemicals. Second, the chemicals selected were those tested for carcinogenicity by the NTP within a specified range of time. This selection procedure should minimize the possibility of chemical selection bias.

The major conclusions of the 73-chemical study (*11*) were: *a*) Individual qualitative concordance of the four STTs with rodent carcinogenicity did not show significant differences among assays (approximately 60%) and were much lower than previous estimates, *b*) there was no complementarity among STTs, and *c*) no battery of tests constructed from these four STTs improved the carcinogenicity predictivity of the SAL assay alone. Initial reaction to these conclusions within the genetic toxicology community was mixed. One of the criticisms was that the 73 chemicals were somewhat atypical and therefore the study needed to be repeated.

To confirm and extend the findings of Tennant et al. (*11*), a follow-up study was conducted for an additional 41 chemicals. The results obtained for the 41 chemicals were similar to those reported for the 73 chemicals (*12*). No significant differences between the two datasets were detected and hence the two datasets were combined into a single dataset of 114. A detailed description of the 114-chemical data set was given in Haseman et al. (*13*). The binary results of the four STTs and rodent carcinogenicity for the 114 chemicals are in Table 3 of Haseman et al. (*13*) and are reproduced here in Table 1 for analysis.

Methods

The results of the four STTs and the rodent carcinogenicity in Table 1 can be summarized in two $2 \times 2 \times 2 \times 2$ contingency tables corresponding to 67 carcinogens and 47 noncarcinogens, respectively. A log-linear analysis can be applied to these two $2 \times 2 \times 2 \times 2$ contingency tables to access the dependence structures of the four STTs. The log-linear model is an analysis of variance (ANOVA)-type data analysis method for a contingency table. It represents the logarithm of expected cell frequency as the sum of main effects of explanatory variables (STTs) and their interactions (*14,15*). The best model is selected based on two competing criteria; parsimony and goodness of fit of the model. The likelihood ratio test of the SAS procedure CATMOD (*16*) can be used to perform the model selection. The SAS procedure CATMOD provides for each model the estimated cell frequencies and their standard errors. Therefore, we can evaluate p(STT results|CA) and p(STT results|NC) for each of the 16 possible STT configurations and each model considered.

Following Rosenkranz, assume that p(CA) = p(NC) = 1/2 for obtaining p(CA|STT results). Application of the Bayes formula permits calculation of the estimates of p(CA|STT results), the probability of carcinogenicity of a chemical given its STT results, as follows:

$$p(CA|STT\ results) = \frac{p(STT\ results|CA)p(CA)}{p(STT\ results|CA)p(CA) + p(STT\ results|NC)p(NC)} = \frac{\hat{m}_{ijkl|CA} / 67}{\hat{m}_{ijkl|CA} / 67 + \hat{m}_{ijkl|NC} / 47} \quad (3)$$

where $\hat{m}_{ijkl\,CA}$ and $\hat{m}_{ijkl\,NC}$ represent estimated expected cell frequencies corresponding to the (i,j,k,l) configuration of the four STTs given carcinogens and noncarcinogens, respectively. The estimated expected cell frequencies provide estimates of p(STT results | CA) and p(STT results | NC) under each possible model. Using this log-linear model approach, one can obtain estimates

Table 1. Binary results of four STTs and rodent carcinogenicity.[a]

STT results				Carcinogenicity results: Original 73 chemicals		New 41 chemicals		Total 114 chemicals	
SAL	ABS	SCE	MLA	+	−	+	−	+	−
+	+	+	+	14	3	8	0	22	3
+	+	+	−	0	0	0	0	0	0
+	+	−	+	2	0	0	0	2	0
+	+	−	−	0	0	0	0	0	0
+	−	+	+	3	1	1	0	4	1
+	−	+	−	0	0	0	0	0	0
+	−	−	+	0	0	2	0	2	0
+	−	−	−	1	0	1	0	2	0
−	+	+	+	5	4	3	2	8	6
−	+	+	−	2	1	0	2	2	3
−	+	−	+	0	0	0	0	0	0
−	+	−	−	1	1	0	0	1	1
−	−	+	+	5	6	2	6	7	12
−	−	+	−	3	1	0	0	3	1
−	−	−	+	2	2	1	4	3	6
−	−	−	−	6	10	5	4	11	14
Total				44	29	23	18	67	47

[a] From Haseman et al. (*13*).

Abbreviations: STT, short-term test; SAL, Salmonella/microsome mutagenicity asssy; ABS, chromosome aberrations; SCE, sister chromatid exchange; MLA, mouse lymphoma L5178Y cell mutagenesis assay.

of p(STT|CA) and p(STT results|NC) under various possible dependence structures. Hence, independence of the four STTs is no longer assumed.

The SAS procedure CATMOD also provides standard errors of the estimated cell frequencies. This allows us to calculate the standard errors of the estimates of p(CA|STT results) as in Equation 4 by using the usual δ method.

$$\text{Var}\left(\hat{p}(\text{CA}|\text{STT results})\right) \approx \frac{m_{CA}^2}{(m_{CA}+m_{NC})^4}\sigma_{CA}^2 + \frac{m_{NC}^2}{(m_{CA}+m_{NC})^4}\sigma_{NC}^2,$$

where

$$m_{CA} = \hat{m}_{ijkl|CA}$$

$$m_{NC} = \hat{m}_{ijkl|NC}$$

$$\sigma_{CA}^2 = [SE(\hat{m}_{ijkl|CA})]^2/(67)^2$$

and $$\sigma_{NC}^2 = [SE(\hat{m}_{ijkl|NC})]^2/(47)^2 \quad (4)$$

The standard errors derived from Equation 4 make the point estimates in Equation 3 more readily interpretable.

Results

Several log-linear models were fit to each of the two 2 × 2 × 2× 2 contingency tables. The goodness of fit of a representative subset of these models is shown in Table 2 in terms of the likelihood ratio test statistic, G^2. Following Fienberg (*14*) the brackets are used to denote a model. In the brackets notation A, C, M, and S stand for ABS, SCE, MLA, and SAL, respectively.

The model [A] [C] [M] [S] is the complete independence model under which CPBS was developed originally. The goodness of fit analysis shows that the four STTs are not conditionally independent. It is concluded here that the model [AC] [CM] [MS] is the best for both carcinogens and noncarcinogens. The expected cell frequencies under each of the above five models are shown in Table 3.

Table 2. Goodness of fit of various models.

	Carcinogens			Noncarcinogens		
Model	G^2	df	p	G^2	df	p
[A] [C] [M] [S]	69.93	11	0.0000	38.64	11	0.0000
[A] [S] [CM]	48.40	10	0.0000	22.67	10	0.0120
[S] [AC] [CM]	29.40	9	0.0006	11.17	9	0.2643
[AC] [CM] [MS]	12.95	8	0.1136	6.77	5	0.2379
[AC] [AS] [CM] [MS]	5.69	7	0.5758	4.08	4	0.3950

Abbreviations: A, chromosome aberrations; C, sister chromatid exchange; M, mouse lymphoma L5178Y cell mutagenesis; S, Salmonella/microsome mutagenicity assay.

Based on the expected cell frequencies and their standard errors (not shown) in Table 3, one can obtain estimates of p(CA|STT results) and their standard errors under various models by using Equations (3) and (4). These estimates and standard errors are shown in Table 4.

Discussion

The observed frequencies for the positive response of SAL and the negative response of MLA (SAL+, MLA−) are 2 for carcinogens and 0 for noncarcinogens, respectively. This suggests that (SAL+, MLA−) is a relatively rare event, for both carcinogens and noncarcinogens. The complete independence model for the carcinogens fails to reveal this characteristic and, in general, shows substantial lack of fit in those cases where the STTs are in complete agreement (Table 3). The model [AC] [CM] [MS] is consistent with this observation; however, it remains to be seen whether a mechanistic interpretation can be attached to it. In relation to the small observed frequencies for

Table 3. Expected cell frequencies under various models.

				Carcinogens						Noncarcinogens					
					Model [a]						Model [a]				
SAL	ABS	SCE	MLA	Observed	1	2	3	4	5	Observed	1	2	3	4	5
+	+	+	+	22	8.2	10.2	13.6	17.8	21.9	3	0.4	0.5	0.9	1.5	2.9
+	+	+	−	0	3.3	1.2	1.7	0.4	0.7	0	0.2	0.1	0.2	0.0	0.0
+	+	−	+	2	3.8	1.7	0.5	0.6	1.1	0	0.3	0.1	0.0	0.0	0.1
+	+	−	−	0	1.5	3.5	1.0	0.2	0.4	0	0.2	0.4	0.1	0.0	0.0
+	−	+	+	4	7.5	9.4	6.0	7.8	4.8	1	1.0	1.4	1.0	1.7	0.7
+	−	+	−	0	3.0	1.1	0.7	0.2	0.1	0	0.6	0.2	0.2	0.0	0.0
+	−	−	+	2	3.4	1.6	2.9	3.8	2.3	0	0.8	0.4	0.5	0.8	0.3
+	−	−	−	2	1.4	3.2	5.7	1.3	0.8	0	0.5	0.9	1.2	0.0	0.0
−	+	+	+	8	9.0	11.2	14.9	10.7	7.3	6	3.9	5.6	9.3	8.7	7.5
−	+	+	−	2	3.6	1.4	1.8	3.1	2.1	3	2.7	1.0	1.7	1.8	1.6
−	+	−	+	0	4.1	1.9	0.5	0.4	0.4	0	3.2	1.5	0.3	0.2	0.2
−	+	−	−	1	1.6	3.8	1.0	1.8	1.2	1	2.1	3.8	0.7	0.7	0.7
−	−	+	+	7	8.2	10.2	6.5	4.7	7.0	12	10.3	14.6	10.8	10.2	10.9
−	−	+	−	3	3.3	1.2	0.8	1.4	2.1	1	7.0	2.6	2.0	2.2	2.4
−	−	−	+	3	3.8	1.7	3.1	2.3	3.3	6	8.3	4.0	5.2	4.9	5.3
−	−	−	−	11	1.5	3.5	6.3	10.7	11.6	14	5.6	9.9	13.1	14.3	14.3
Total				67						47					

Abbreviations: SAL, Salmonella/microsome mutagenicity assay (S); ABS, chromosome aberrations (A); SCE, sister chromatid exchange (C); MLA, mouse lymphoma L5178Y cell mutagenicity assay (M).

[a] Model 1 = [A] [C] [M] [S], model 2 = [A] [S] [CM], model 3 = [S] [AC] [CM], model 4 = [AC] [CM] [MS], model 5 = [AC] [AS] [CM] [MS].

Table 4. Estimates of p(CA|STT results) and their standard errors. [a]

SAL	ABS	SCE	MLA	Model: Independence	[A] [S] [CM]	[S] [AC] [CM]	[AC] [CM] [MS]
+	+	+	+	0.935 (0.1828)	0.933 (0.1728)	0.917 (0.1581)	0.893 (0.1426)
+	+	+	−	0.920 (0.2223)	0.903 (0.3783)	0.881 (0.3568)	1.000 (0.8014)
+	+	−	+	0.899 (0.2218)	0.897 (0.3197)	0.993 (0.5706)	1.000 (0.5460)
+	+	−	−	0.840 (0.2235)	0.874 (0.2249)	0.917 (0.5032)	1.000 (0.8867)
+	−	+	+	0.840 (0.1532)	0.829 (0.1418)	0.806 (0.1795)	0.763 (0.1588)
+	−	+	−	0.778 (0.1646)	0.765 (0.2755)	0.736 (0.2757)	1.000 (0.8269)
+	−	−	+	0.749 (0.1563)	0.752 (0.2299)	0.805 (0.2541)	0.769 (0.2269)
+	−	−	−	0.663 (0.1502)	0.708 (0.1564)	0.768 (0.1699)	1.000 (0.7154)
−	+	+	+	0.618 (0.0878)	0.585 (0.0820)	0.529 (0.0778)	0.463 (0.0093)
−	+	+	−	0.483 (0.1070)	0.486 (0.1785)	0.430 (0.1903)	0.547 (0.1714)
−	+	−	+	0.473 (0.1039)	0.469 (0.1538)	0.584 (0.2878)	0.584 (0.3031)
−	+	−	−	0.348 (0.1497)	0.414 (0.1209)	0.529 (0.2778)	0.643 (0.2735)
−	−	+	+	0.358 (0.0879)	0.330 (0.0859)	0.297 (0.1226)	0.244 (0.1279)
−	−	+	−	0.249 (0.1383)	0.248 (0.2776)	0.221 (0.3129)	0.309 (0.2498)
−	−	−	+	0.243 (0.1308)	0.236 (0.2314)	0.296 (0.1947)	0.248 (0.2256)
−	−	−	−	0.158 (0.1854)	0.198 (0.1520)	0.252 (0.1261)	0.344 (0.0989)

Abbreviations: SAL, Salmonella/microsome mutagenicity assay (S); ABS, chromosome aberrations (A); SCE, sister chromatid exchange (C); MLA, mouse lymphoma L5178Y cell mutagenicity assay (M).

[a] Standard errors are in parenthesis.

(SAL+, MLA−), note in Table 4 large standard errors of the estimates of p(CA|STT results), particularly for (SAL+, MLA−) under the [AC] [CM] [MS] model.

Earlier findings (*11–13*) suggest that collapsing the four-dimensional contingency table into a lower-dimensional contingency table does not improve the predictivities of carcinogenicity in Table 4. The stepwise model improvement in Table 2 is reminiscent of stepwise regression. Model improvement over the complete independence model first brings in [CM], the interaction of SCE and MLA. Then [AC], the interaction of ABS and SCE enters into the model. It is noticeable that SAL is weakly linked only at the third step as [MS] enters the model. Table 4 shows how the estimates of p(CA|STT results) change under various depencence structures of the four STTs. dependence structures of the four STTs. Most notable are the generally large standard errors associated with the predicted probabilities of carcinogenicity. Even a database of 114 chemicals is still relatively small for this qualitative prediction. Investigators interested in applying this methodology to analysis of their own data, proprietary or otherwise, need to appreciate the relationship between precision of the prediction and the size of the available database. The standard errors of the estimates serve this purpose.

Finally, why should biological end points as diverse as the four STTs considered here show dependence? Is there a common precursor, such as genomic fluidity? Irrespective of the explanation for the dependence, Kuroki and Matsushima (*17*) were correct when they wrote,

> Although there have been a number of studies on the correlation of results of short-term tests with carcinogenicity, not much attention has been paid to the correlation between results of short-term tests. However, this is important in evaluating short-term tests and determining complementary assays that in combination may increase predictive values. Such correlations may suggest cellular and molecular mechanisms by which a given chemical causes cancer.

REFERENCES

1. Heinze, J. E., and Poulsen, N. K. The optimal design of batteries of short-term tests for detecting carcinogens. Mutat. Res. 117: 259–269 (1983).
2. Rosenkranz, H. S., Klopman, G., Chankong, V., Pet-Edwards, J., and Haimes, Y. Y. Prediction of environmental carcinogens: a strategy for the mid-1980s. Environ. Mutagen. 6: 231–258 (1984).
3. Chankong, V., Haimes, Y. Y., Rosenkranz, H. S., and Pet-Edwards, J. The carcinogenicity prediction and battery selection (CPBS) method: a Bayesian approach. Mutat. Res. 153: 135–166 (1985).
4. Pet-Edwards, J., Chankong, V., Rosenkranz, H. S., and Haimes, Y. Y. Application of the carcinogenicity prediction and battery selection (CPBS) method to the Gene Tox data base. Mutat. Res. 153: 187–200 (1985).
5. Ennever, F. K., and Rosenkranz, H. S. Short-term test results for NTP non-carcinogens: an alternate, more precise battery. Environ. Mutagen. 8: 849–865 (1986).
6. Ennever, F. K., and Rosenkranz, H. S. Application of the carcinogenicity prediction and battery selection method to recent National Toxicology Program short-term test data. Environ. Mol. Mutagen. 13: 332–338 (1989).
7. Denniston, C. Screening program theory: decision functions and protocols. Environ. Mutagen. 7: 53–72 (1985).
8. Lave, L. B., and Omenn, G. S. Cost effectiveness of short-term tests for carcinogenicity. Nature 324: 29–34 (1986).
9. Ennever, F. K., and Rosenkranz, H. S. Prediction of carcinogenic potency by short-term genotoxicity tests. Mutagenesis 2: 39–44 (1987).
10. Purchase, I. F. H., Longstaff, E., Ashby J., Styles, J. A., Anderson, D., Lefevre, P. A., and Westwood, F. R. An evaluation of 6 short-term tests for detecting organic chemical carcinogens. Br. J. Cancer. 37: 873–959 (1978).
11. Tennant, R. W., Margolin, B. H., Shelby, M. D., Zeiger, E., Haseman, J. K., Spalding, J., Caspary, W., Resnick, M., Stasiewicz, S., Anderson, B., and Minor, R. Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. Science 236: 933–941 (1987).
12. Zeiger, E., Haseman, J. K., Shelby, M. D., Margolin, B. H., and Tennant, R. W. Evaluation of four in vitro genetic toxicity tests for predicting rodent carcinogenicity: confirmation of earlier results with 41 additional chemicals. Environ. Mol. Mutagen. 16:(suppl. 18): 1–14 (1990).
13. Haseman, J. K., Zeiger, E., Shelby, M. D., Margolin, B. H. and Tennant, R. W. Predicting rodent carcinogenicity from four in vitro genetic toxicity assays: an evaluation of 114 chemicals studied by the National Toxicology Program. J. Am. Stat. Assoc. 85: 964–971 (1990)
14. Fienberg, S. E., The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge, MA, 1980.
15. Fingleton, B. Models of Category Counts. Cambridge University Press, Cambridge, 1984.
16. SAS Institute Inc. SAS Users Guide: Statistics, Version 5 ed. Cary, NC, 1985.
17. Kuroki, T., and Matsushima, T. Performance of short-term tests for detection of human carcinogens. Mutagenesis 2: 33–37 (1987).